Semantics and Ontologies For Geospatial Architecture: A Manifesto of Ideas

Presented by Dr. Gary Berg-Cross

Based on the work of (alphabetically)

Gary Berg-Cross, Isabel Cruz, Mike Dean, Tim Finin, Mark Gahegan, Pascal Hitzler, Hook Hua, Krzysztof Janowicz, Naicong Li, Philip Murphy, Bryce Nordgren, Leo Obrst, Mark Schildhauer, Amit Sheth, Krishna Sinha, Anne Thessen, Nancy Wiegand, and Ilya Zaslavsky

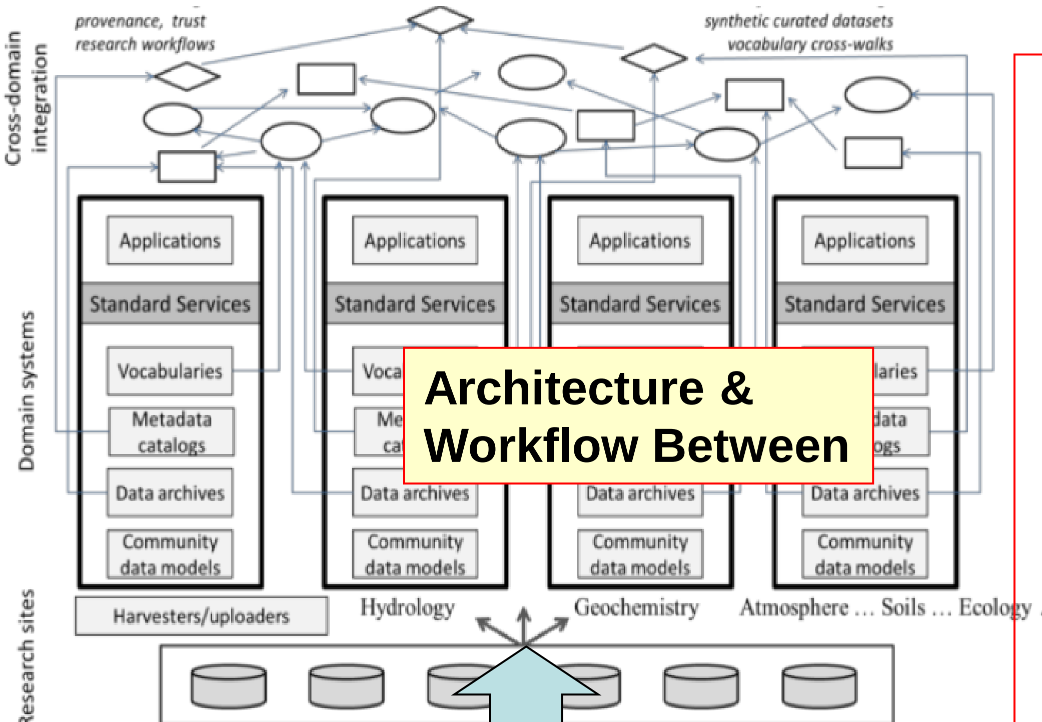
Workshop on Semantics in Geospatial Architectures: Applications and Implementation

October 28-29, 2013

Pyle Center (702 Langdon Street, Madison, WI), University of Wisconsin-Madison

Paper at http://stko.geog.ucsb.edu/gibda2012/

Graphic Overview of S/O (EarthCube) Manifesto



Architecture & Workflow Between

"Insertion"

Guiding principles

1. Uses Cases
2. Lightweight -opportunistic
3. Semantic interoperability with semantic heterogeneity
4. Bottom-up & top-down approaches
5. Domain - ontology engineer teams
6. Formalized bodies of knowledge across Earth science domains
7. Reasoning services

Knowledge Infrastructure **Vision**

Community Understanding of Semantic role and value

Manifesto Presentation Overview

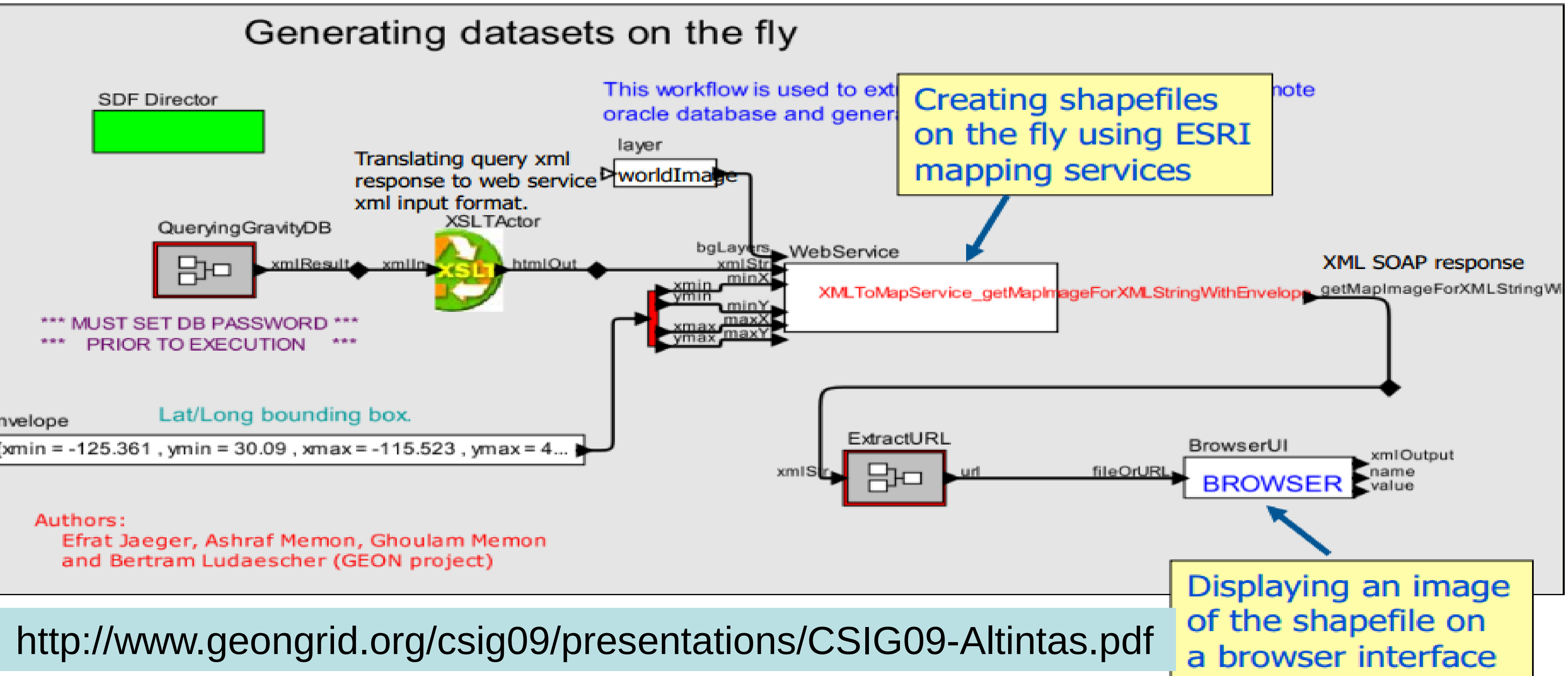
1. Infrastructure & Semantic Architecture Background
    1. Goals
    2. Workflow & Mediation
    3. LOD - a driver
2. Fostering Understanding of a Vision
    1. Next generation visions and role in generic **“knowledge infrastructure”**
    2. Communicate **value proposition** of semantic technologies (in non-technical language).
3. Guiding Methodological Principles for Success
    1. Use Cases
    2. Lightweight -opportunistic conceptual, formalization efforts
    3. Semantic interoperability that protects semantic heterogeneity
    4. Bottom-up and top-down semantics approaches
    5. Integrated ontological engineering team
    6. Formalized bodies of knowledge across Earth science domains
    7. Reasoning services

10 Year Infrastructure Goals/Strategy

- *Fill a Need to* ***collaboratively create a community, knowledge management system*** *and infrastructure/cyberinfrastructure*
    1. ***converges*** *on and* ***integrates*** *important(****BIG****) geosciences data in an open, transparent and inclusive manner.*
    2. ***Something easily adopted*** *by geosciences researchers & educators.*
    3. ***exposes*** *data and information to knowledge creation through data-enabled science*
    4. *Enhance* ***Interworkability*** *of data and information (****shared workflows****)*
- *Strategy*
    1. *introduce* ***new approaches and technologies*** *(****SEMANTIC TECH****) and/or combining productive* ***tools*** *and solutions in different ways.*
    2. *promote integration, flexibility, inclusiveness, and easy adoption by* ***connecting the several layers of data and information management****, from the resource layer with access to data and information, to the data curation and management layer.*

Example of Semantic Technologies & Modern Infrastructure

- Increasing role formalizing scientific workflow
  - DB access & querying steps, data analysis & mining steps etc.

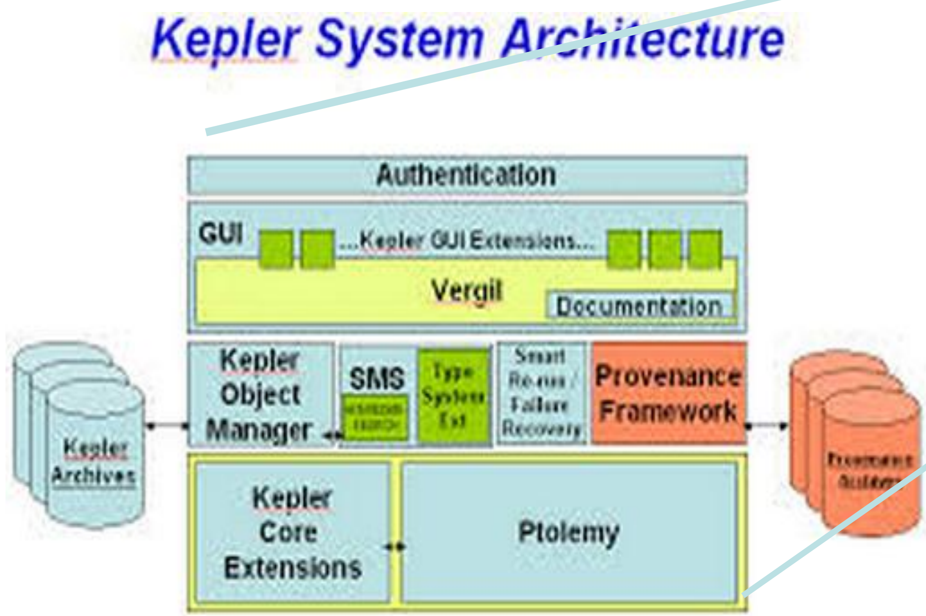


http://www.geongrid.org/csig09/presentations/CSIG09-Altintas.pdf

Kepler & 3-Tiered GEON Portal & GRID

Provides experience integrating heterogeneous local & remote tools in 1 interface

- Web, Grid & GIS services are formalized a bit
- Relational and spatial databases access
- Reusable generic and domain specific actors… etc.

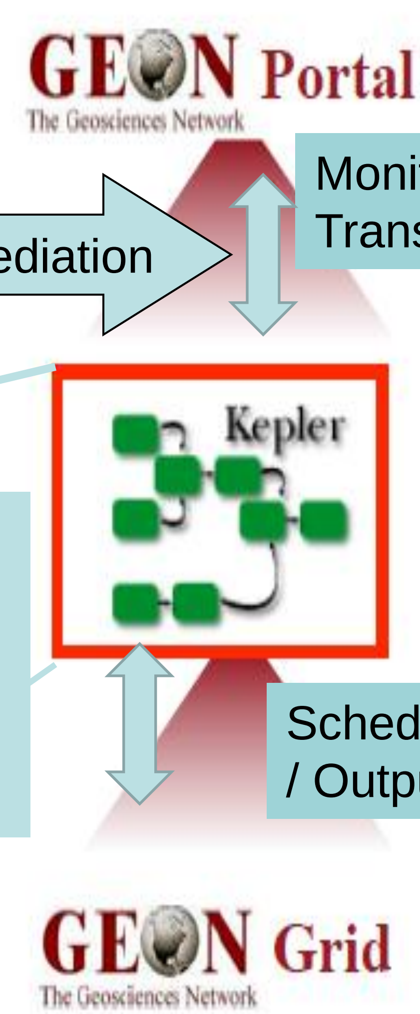


Semantic Mediation

Monitoring/ Translation

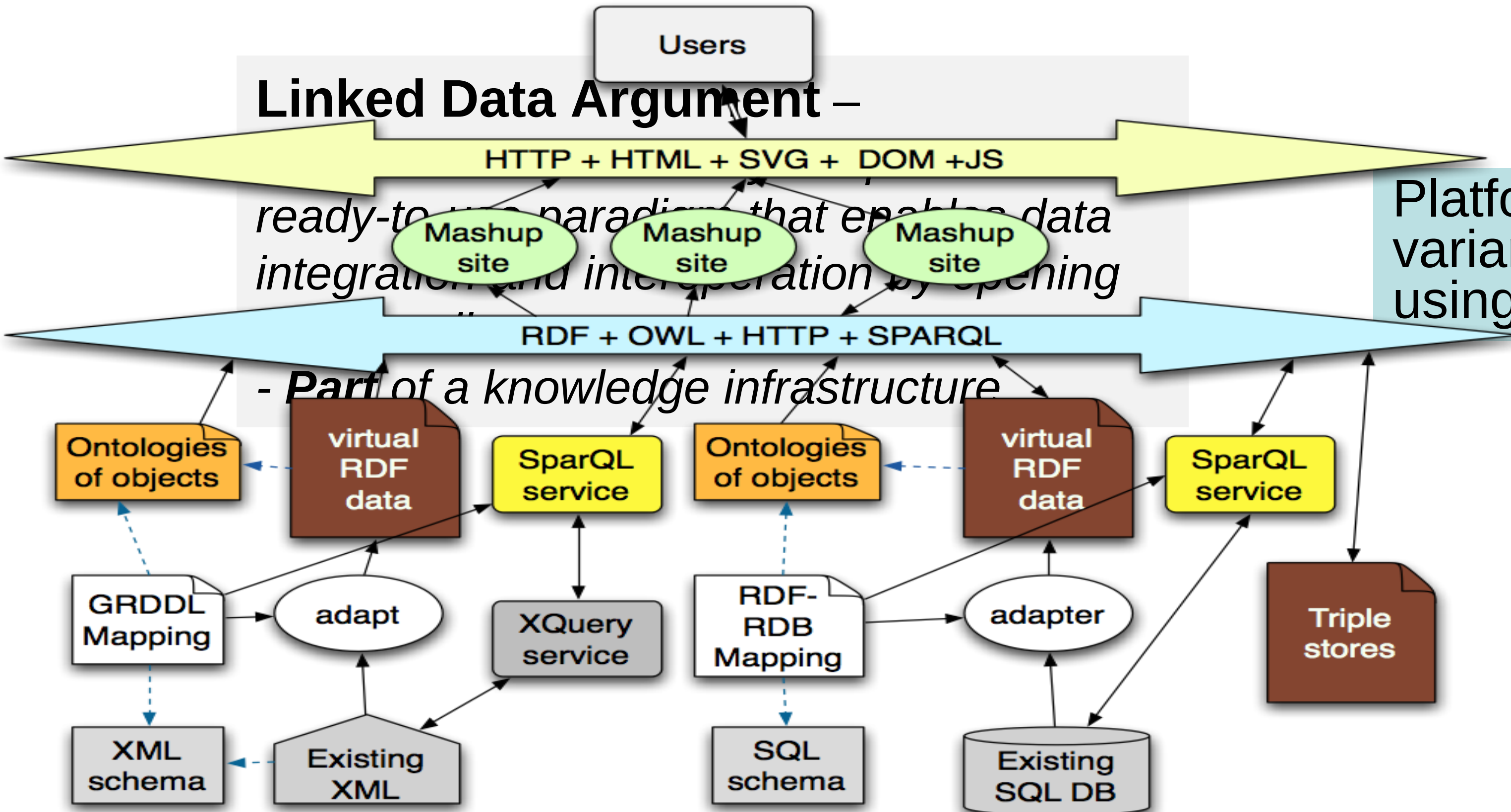
Knowledge-based infrastructures for **semantic annotation** of metadata
Supports Search

Scheduling / Output



Driver: Semantic Web & Linked Data

- **Many** Semantic Tech parts but:
  - an important driver has been the Semantic Web & Linked Open Data (LOD) framework

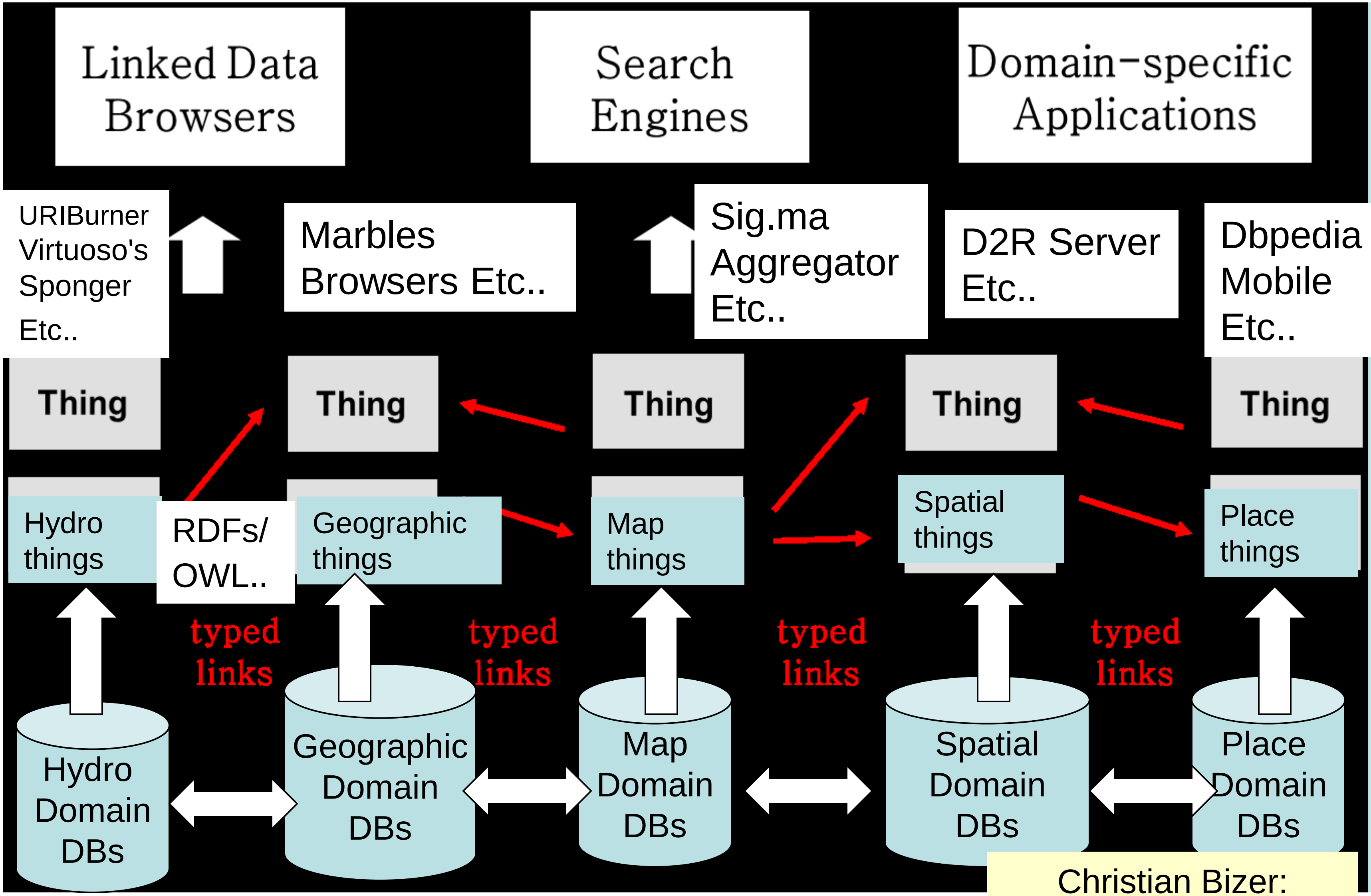


Platform agnostic variant of ODBC etc. using hyperlinks

Ontologies & KR languages for intended meaning

[Semantics in Geospatial Architectures]

Linked Data, Applications, Horizontal & Vertical Integration

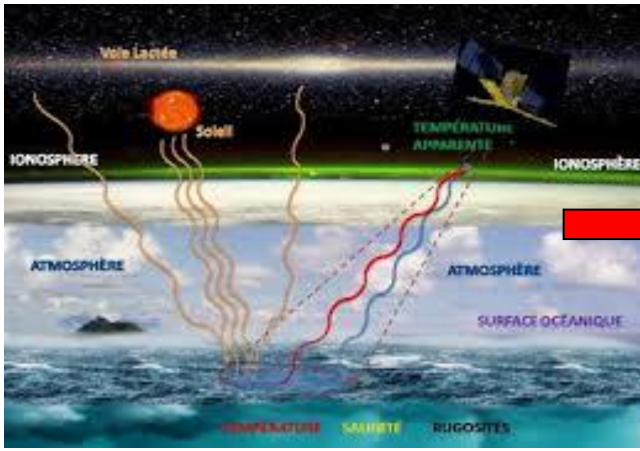


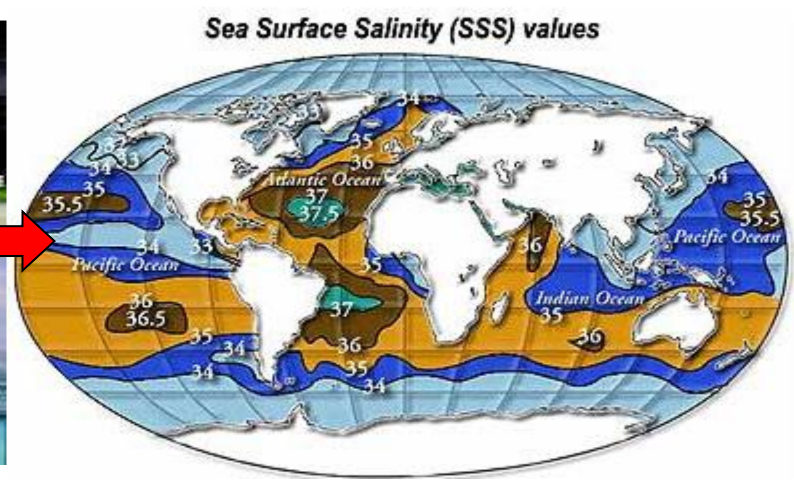
Semantics in Geospatial Architectures

Christian Bizer: The Web of Linked Data (26/07/2009)

Do data integration, analysis, & visualization steps "Behind the Scenes"

Problem Semantic technologies **require knowledge of formal logic** that is unfamiliar to most Earth scientists. So **Institutionalize** what we can.

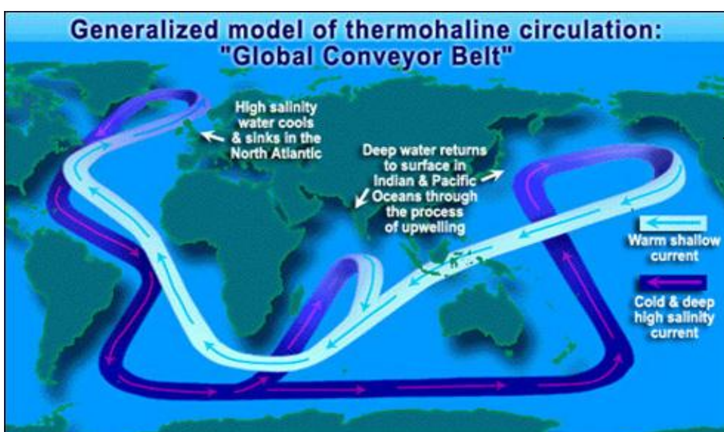




chemical concentrations

sea surface temperature

Observations of sea surface temp (SST) & salinity measurements from the sea surface at a location



Model-data relation

Automatically link the *data via terms and Correlated measurements from locations situated near to location*.

"You mean I don't have to be able to read XML. RDF or OWL? Yea!!!

Communicating an Understandable Value Proposition

- What is proposed?
  - **Uncover hidden heterogeneities** & make them explicit
    - This affords key incompatibility discovery, prevent users from mixing apples & oranges
- How:
  - Promote common vocabularies for annotating and describing data using terms in formalized ontologies
  - Leverage vast number of available repositories, ontologies, methods, standards, and tools that support scientists in publishing, sharing, and discovering data
- Value > expected from annotation using simple metadata

- *But the community needs to **understand the semantic technologies vision-infrastructure-value in a non-technical language……and believe that this can be done without heroic efforts.***

Seven (or so) Guiding Principles for Facilitating Implementation and Application

Methods

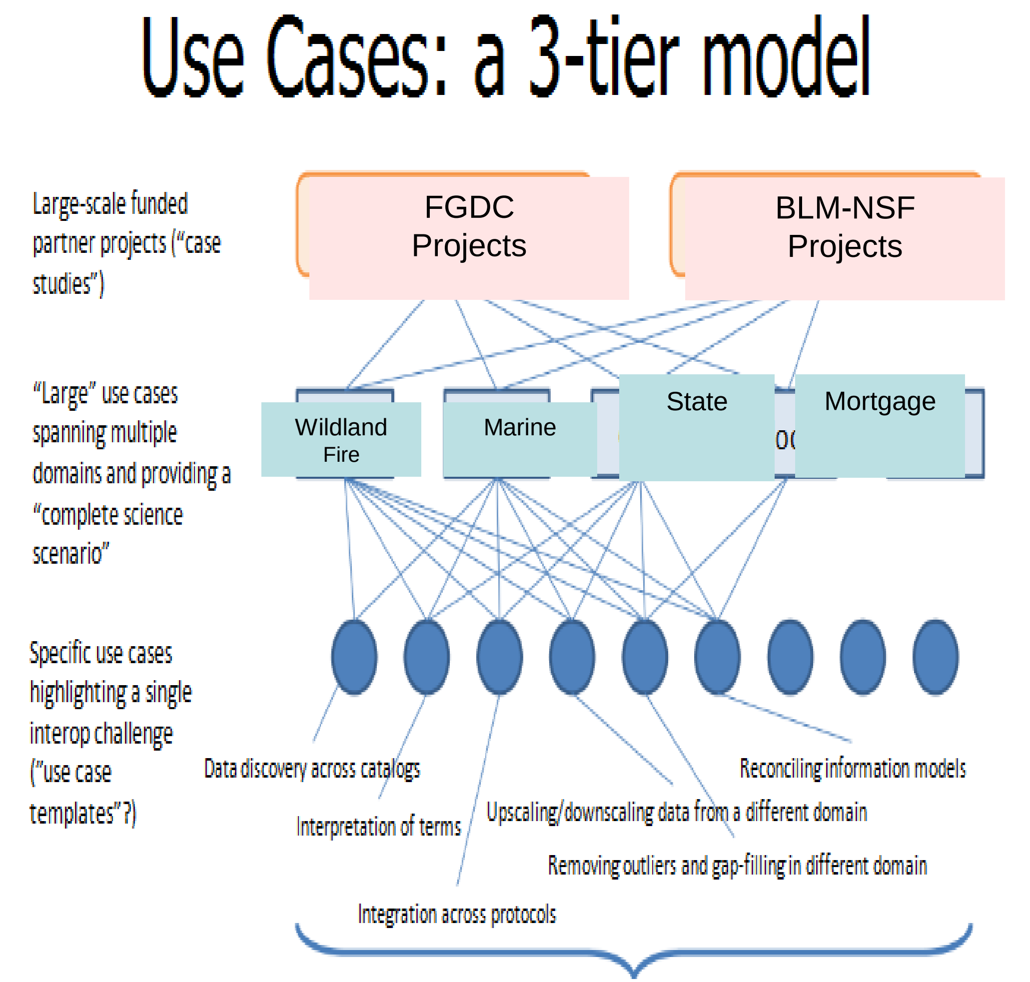
1. Driven by **concrete use cases** and GIScience/ practitioner needs
2. Use lightweight (semantic) approaches
3. Foster semantic interoperability without restricting extant semantic heterogeneity
4. Employ bottom-up AND top-down semantics approaches
5. Involve & enable domain experts assisted by ontology engineers
6. Use S &O to build a formal body of knowledge in various GIscience domains

Technology

7. Employ classical and non-classical reasoning services

1. Understand Requirements: Concrete Use Cases

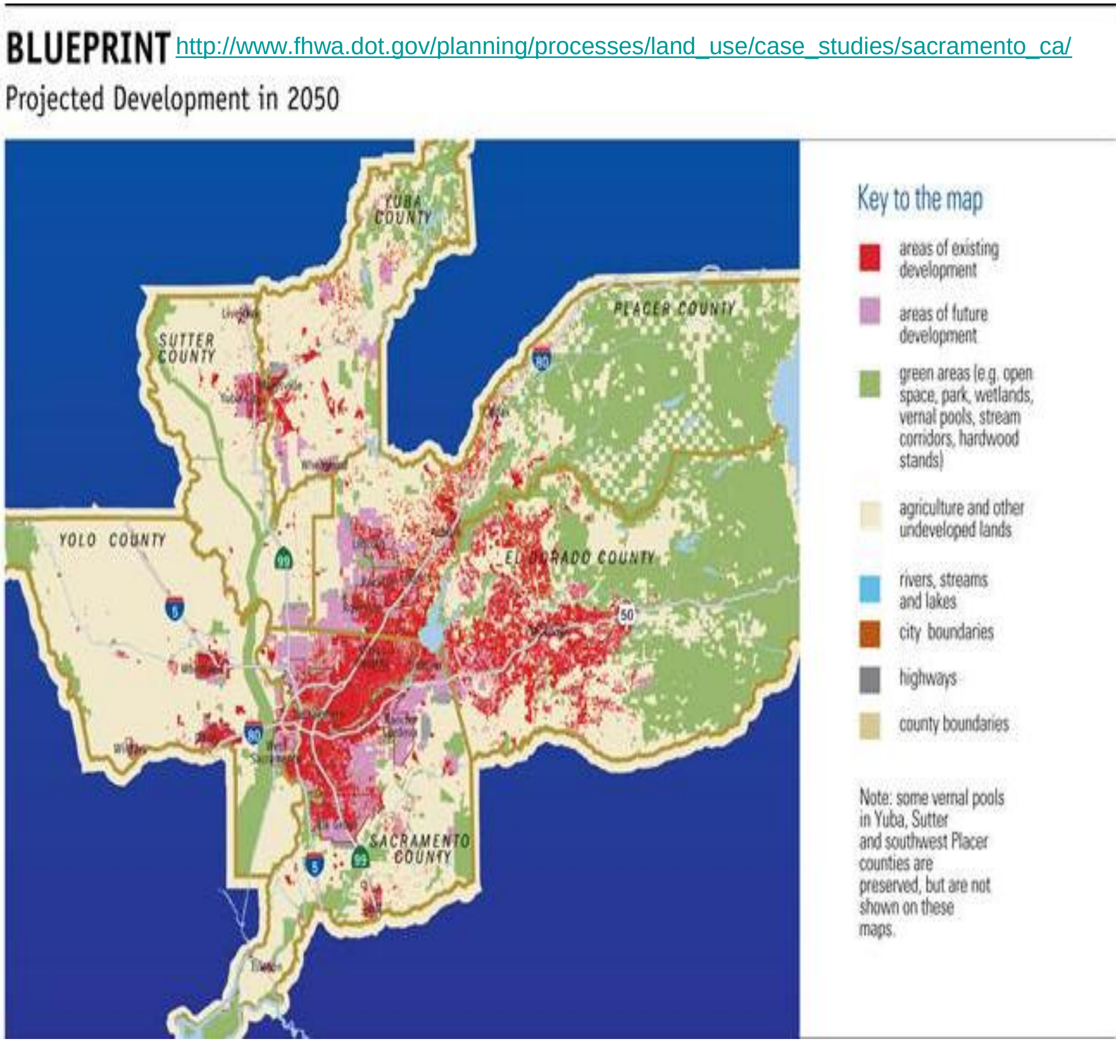
- Work should be driven by use cases generated by members of the GS community – e.g. Land Parcels/cadastral?
- Need a substantial study of **interconnected use cases** which expose requirements related to data, models, and tools
  - which have clear implications for data interoperability, ontology, and semantics infrastructure



Notional State/County/City Planning using Land Parcels

- Large area for planning integrating community info, urban planning and design, etc.
- Inputs a range of zoning designations to each land parcel in a given area
- Requires integration of data from several sources of different types
  - Improved parcels models to allow this integration

http://www.fhwa.dot.gov/planning/processes/land_use/case_studies/sacramento_ca/

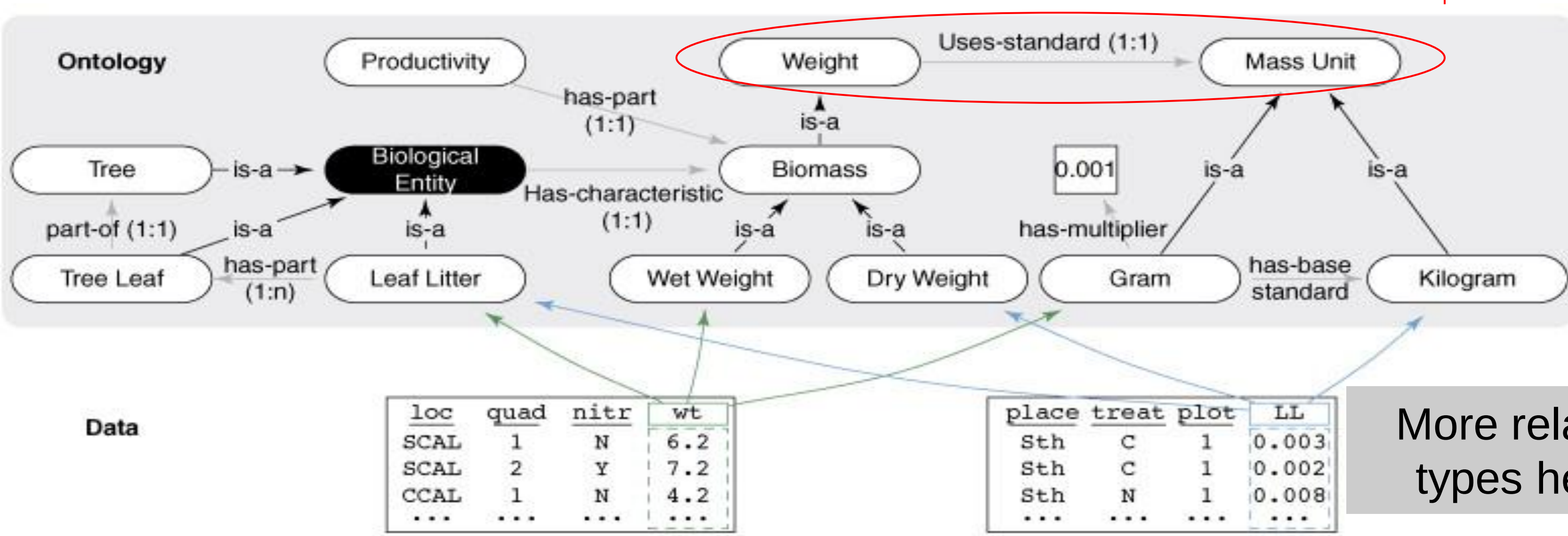


2. Lightweight Methods & Products

- Choose lightweight approaches to support application needs and **reduced entry barrier**
- Low hanging fruit **leverages initial vocabularies** & existing **conceptual models** to ensure that a semantics-driven infrastructure is available for **early use.**

Simple parts/patterns & direct relations to data

Triple like parts



More relation types here.

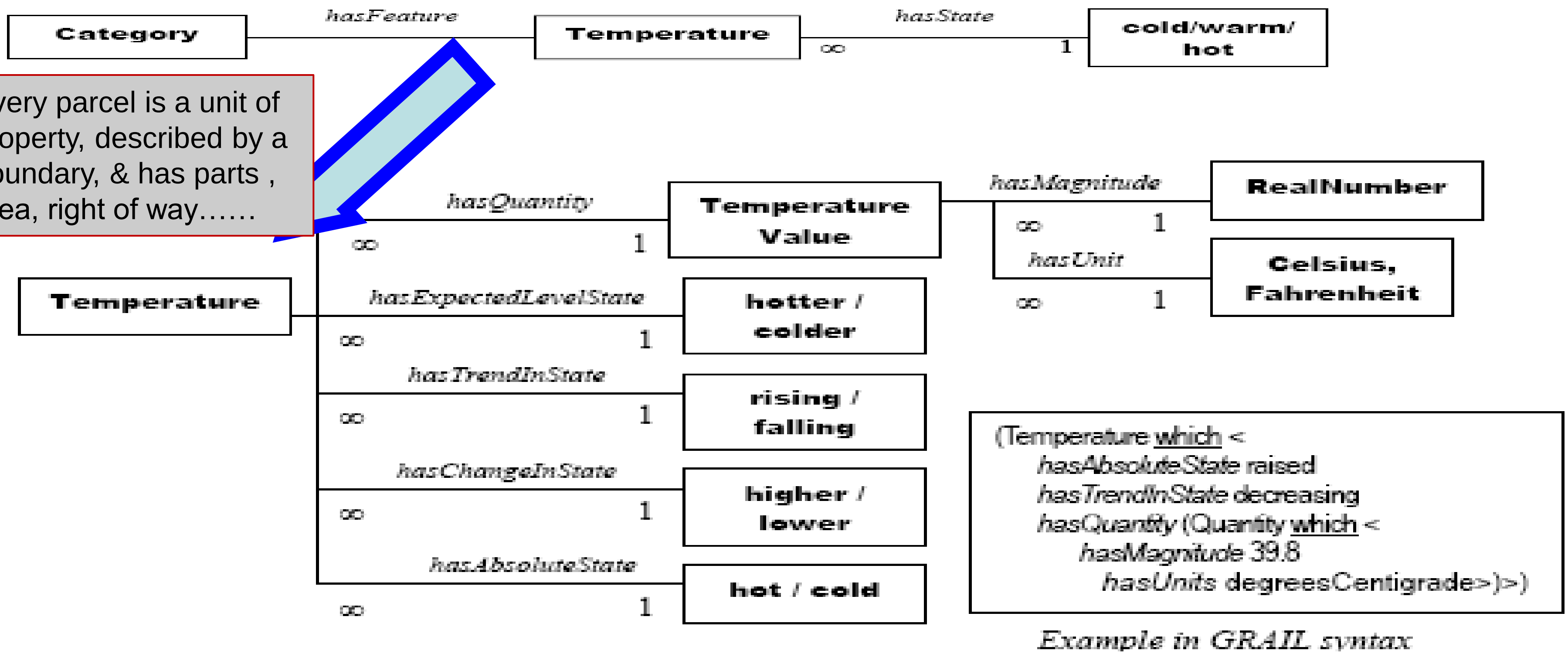
Incremental Approaches: Richer Schemata & Reusable Patterns

Land Parcel, owner….

area, boundary, encumberance….

19 sq ‘, located at.

Every parcel is a unit of property, described by a boundary, & has parts , area, right of way……



Simple Feature-State Model (from GRAIL) becomes a richer schema

Adding Better Semantic Relations/Properties

Kate Beard's point - **Irreflexive, Anti-symmetric** & **Transitive** constructs that captures common understanding.

Observation –Streams flow into rivers etc.

- Property "flows-into" is irreflexive
  - any one river or stream cannot flow into itself as a loop
- "flows-into" is also anti-symmetric
  - if one river flows into the second, the second one can't flow into the first.
- Transitive property for Regions means that the subRegionOf property between Regions is transitive
  - <owl:TransitiveProperty rdf:ID="subRegionOf"> <rdfs:domain rdf:resource="#Region"/> <rdfs:range rdf:resource="#Region"/> </owl:TransitiveProperty>

If Madison, Dane County and WI are regions, and Madison is a subRegion of Dane County , Dan County is a subRegion of WI , then Madison is also a subRegion of WI.

Organizing Relations - Three Kinds of "Structure"

Meronymic	Spatial	Properties
has-part	is-at	[illegible]
has-region	is-inside	[illegible]ter
material	is-outside	height
possesses	abuts	area
element	is-between	depth
	is-along	volume

Relations in GeoSPARQL

Enable **reasoning services**

fishing zone has-depth with average value x

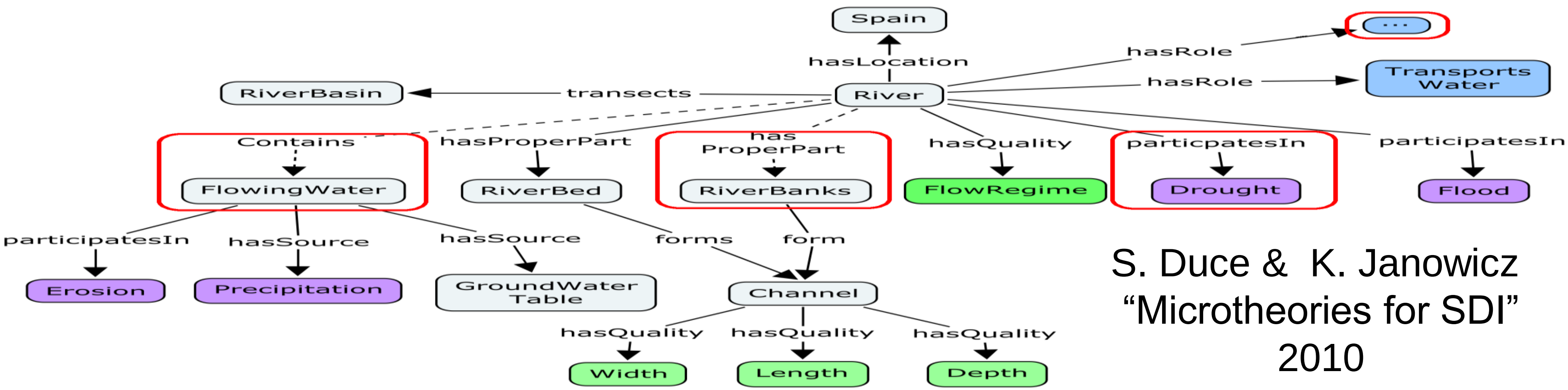
Gulf of Mexico **has-part** gulf fishing zone which has-volume y which **is-inside Gulf** pollution zone
Zone A has area Z……...is-inside Gulf…..has-constituent-nitrogen

3. Foster Semantic Interoperability without restricting underlying Semantic Heterogeneity

Problem: Heterogeneity is introduced by the diverse communities using geospatial concepts.

Solution: Provide **methods** that enable users to flexibly load and **combine** different ontologies instead of hardwiring data to particular ontologies and, thus, hinder their flexible reusability.

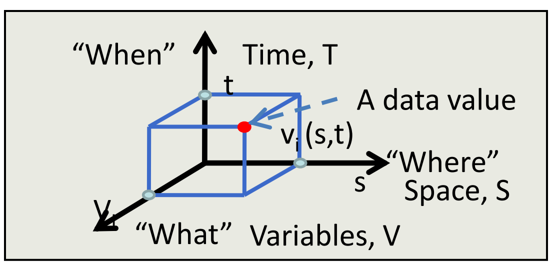
- Example - Work from modular building blocks with microtheories of locally valid semantics
  - Manage multiple, small internally consistent ontologies and focus on interrelations as needed for inter-operation.

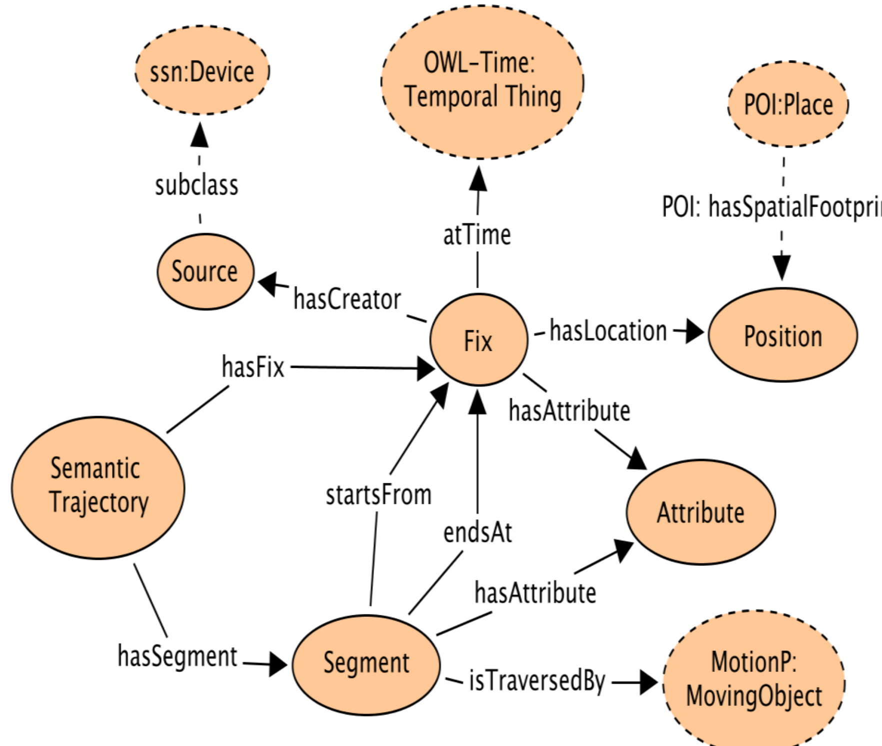


S. Duce & K. Janowicz
"Microtheories for SDI"
2010

Useful Schema - Content Ontology Design Patterns (ODPs) –Semantic Trajectory Pattern Example

- ODPs (aka microtheories) small, modular, & coherent schemas like Temperature.
- Relatively autonomous but conceivably composable with other schemas.
  - E.g. Trajectories/spatial paths, Point Of Interest (POI)- observation area.
- Semantic Trajectory example
    - Indexed by Space-Time-Variable dimensions
  - When we annotate path points of interest (aka Fix) & object motion it is called a Semantic Trajectory
  - Can be bottom up- data driven





ODPs developed at GeoVoCampSB2012 & DaytonGeoVocamp2012

Zhixian Yan. Towards Semantic Trajectory Data Analysis: A Conceptual and Computational Approach. VLDB 2009.

Selfish Plug for Upcoming Workshop

I to want mention the free annual SOCoP Workshop – a **GeoVoCamp**

Ballston VA at the NSF facility
on **Nov 18-19** (M-T) 2013

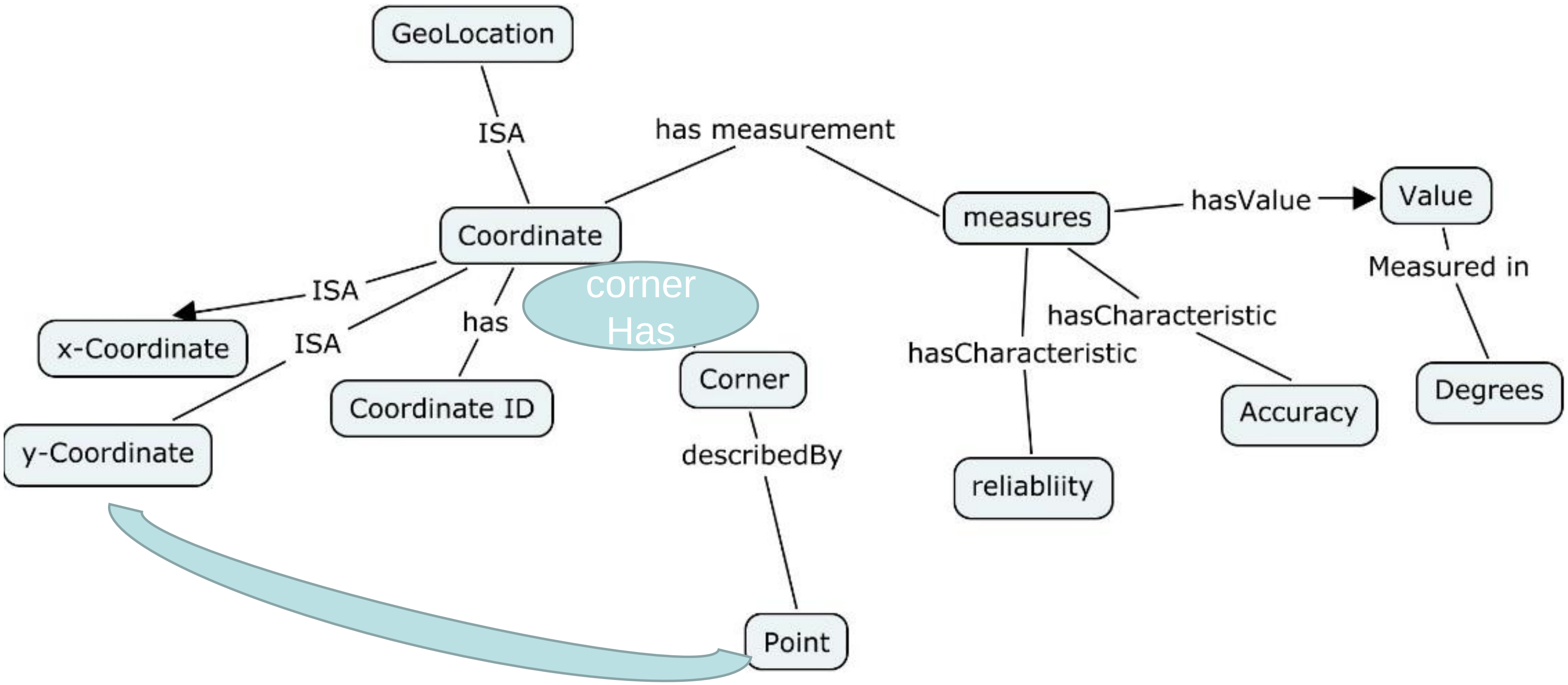
As with previous workshops this will be organized around 3-4 Work Groups. :

- **“Surface Water” - how water sits in terrain**. This is a continuation of last year's (GeoVoCampDC2012) terrain and surface network concepts work
- **Green Building Architecture** (see Charles Vardeman)
- **Ontology patterns to help semantic annotation of maps**

Follow-up to prior GeoVoCamps including those held in Santa Barbara, Dayton and DC in 2012 and at Santa Barbara CA in 2013.

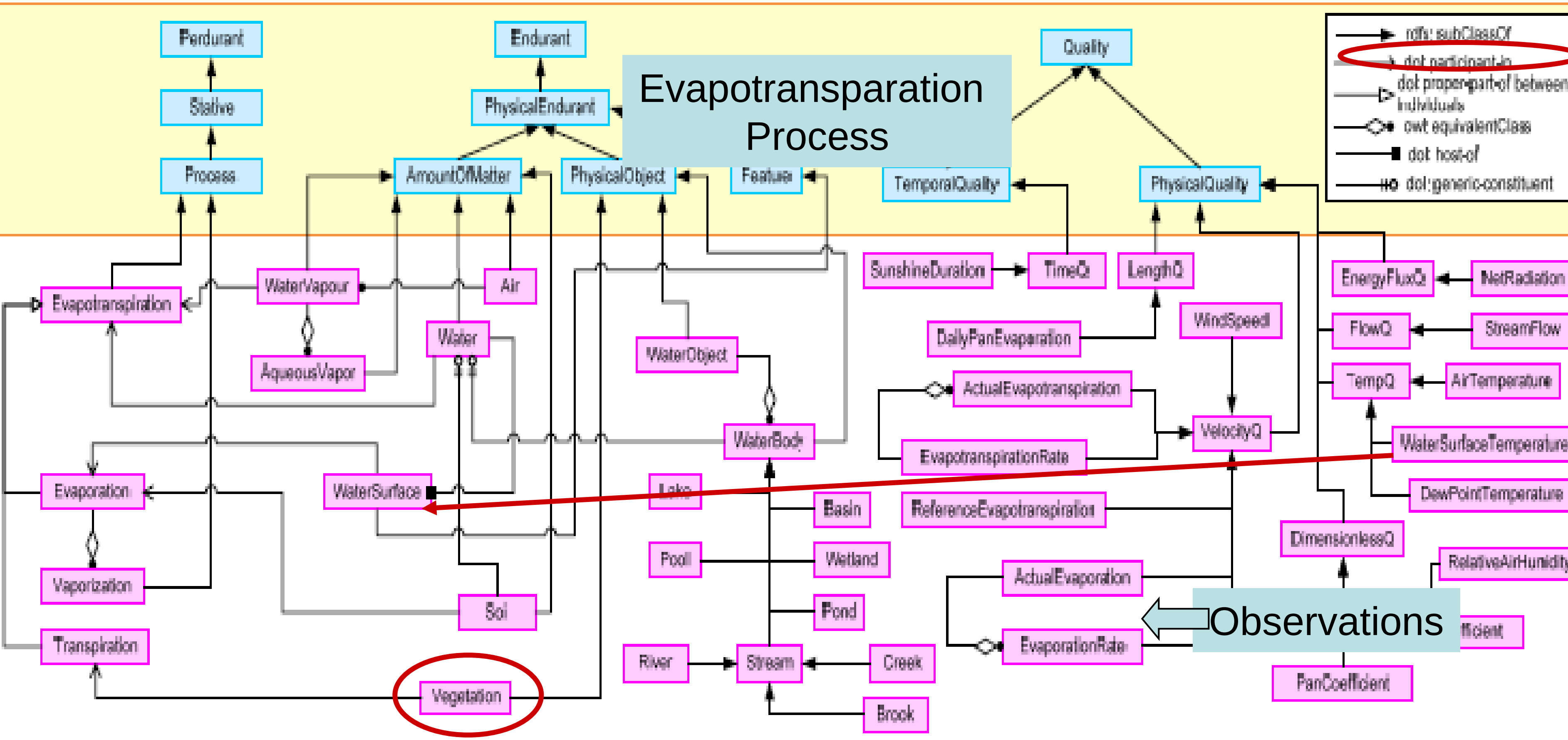
See http://**vocamp.org/wiki/GeoVoCampDC2013**

Notional Example of Corner Pattern



4. Allow for Bottom-up & Top-down Approaches to Semantics

This will ensure a vertical integration from the observations-based data level up to the theory-driven formalization of key domain facts.



Devaraju and Kuhn 2010 developed a design pattern for evaporation as part of a Hydrology domain and mapped it to DOLCE.

5. Integrated KE Teams & Process- domain experts and semantic technologists

- Projects must be structured so domain **experts are active participants** in building semantic models from use cases thru conceptualization to validating final products
- Use :
  - Consistent strategies & methods,
  - Facilitate good documentation, and
- We need **Educational Workshops** on how to do this and also publish, retrieve, and integrate data, models, and workflows.

6. Methods for Useful, Formalized Bodies of Knowledge (10 year goal?)

- Apply ontological engineering/KE to capture the **body of knowledge** for various GI related domains:
  - Conceptualization of **local** models,
  - Work on primitives, i.e., base symbols, for such ontologies,
  - **Ground** primitives in real observations and **align** them to knowledge patterns,
  - **Track** categorical data back to measurements using provenance
    - (e.g. RDF in context),
  - Work to make ontologies first class citizens **usable** by statistical methods.
  - After construction phase, organize building blocks & ontological models
    - To help access data, domain models and their use in tools,
    - This can also be used for **educational applications** for learning about domain concepts, and extracting information

7. Provide Reasoning Services for Products Developed by our Methods

- Behind the scenes - classical and non-classical reasoning services leveraging resources for :
  - organizing and accessing data,
  - models and tools,
  - learning about them, and
  - extracting information
- Reasoning services can be used to :
  - Develop friendly user interfaces,
  - **Dialog systems**
  - Scientist assisting/associate services (chains) for
    - discovering data
    - integrity constraint checking
    - generation of new knowledge and hypothesis testing.

Roadmap for Next Generation Vision

- Use Semantic Web for vertical and horizontal integration
  – centrally important to SDI
- Proposal to redefine Digital Earth as a **knowledge engine*** to support scientists with more than data retrieval.
  – IBM's DeepQA architecture & Semantic Web/Linked Data progress
  – “Reasoning” support is an important addition

* Janowicz, K., Hitzler, P.: The Digital Earth as knowledge engine. Semantic Web Journal

Closing Remarks and Comments

- ***While many details need to be added these should come from continued dialog such as afforded by:***
  - ***VoCAMPs (Vocamp.org)***
  - **Ontolog Mini-Series,**
  - **and other hands on workshops such as SOCoPs annual one in DC**
    - ***Next one is Nov 18-19 at NSF.***
    - http://vocamp.org/wiki/GeoVoCampDC2013

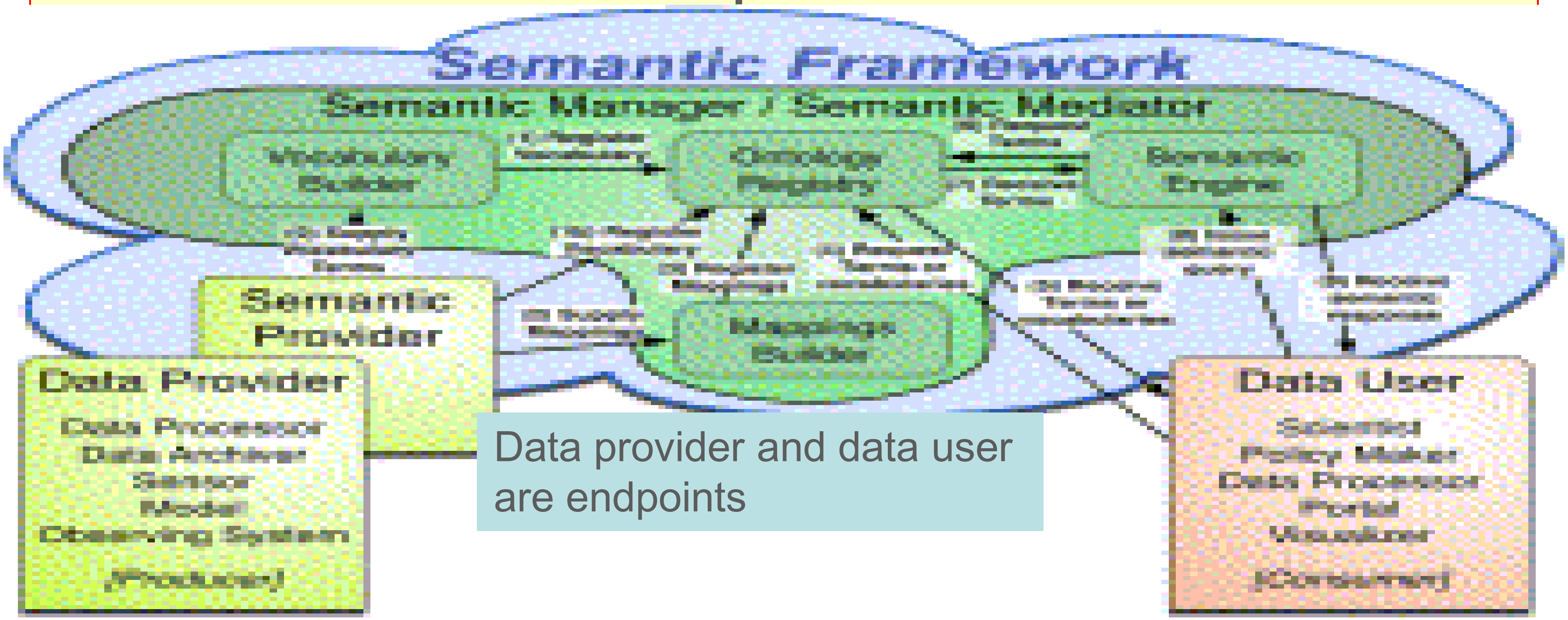
Thank You…. Questions?

Some References & Links

- Hitzler, P., Janowicz, K., Berg-Cross, G., Obrst, L., Sheth, A., Finin, T., Cruz, I.: Semantic Aspects of EarthCube. Technical report, Semantics and Ontology Technical Committee. (2012)
  - http://knoesis.wright.edu/faculty/pascal/pub/EC-SO-TC-Report-V1.0.pdf
- Managing Scientific Data: From Data Integration to Scientific Workflows
- http://users.sdsc.edu/~ludaesch/Paper/gsa-sms.pdf (Ludascher et al.)
- Janowicz, K., Hitzler, P.: The Digital Earth as knowledge engine. Semantic Web Journal 3(3) (2012) 213–221
- EarthCube http://www.nsf.gov/geo/earthcube/ and
  - the community page at http://earthcube.ning.com/
- http://vocamp.org/wiki/GeoVoCampDayton2012
- Earth-Science-Ontolog Mini-Series
  - http://ontolog.cim3.net/cgi-bin/wiki.pl?EarthScienceOntolog
- Kepler See http://www.geongrid.org/csig09/presentations/CSIG09-Altintas.pdf
- S. Duce & K. Janowicz “Microtheories for Spatial Data Infrastructures” https://geog.ucsb.edu/~jano/duce_janowicz_microtheories_giscience2010.pdf
- Christian Bizer: The Web of Linked Data (26/07/2009)

  http://en.wikipedia.org/wiki/Linked_Data Source:

Semantic Framework with major Architecture Components



Data provider and data user are endpoints

Semantic mediator provides the capabilities to link or associate the vocabulary terms found within the semantic manager layer.

Semantic mediation of vocabularies for ocean observing systems, Graybeal et al, 2012